P&S Mobile Genomics

Lecture 3:
Introduction to Sequencing

Dr. Mohammed Alser
@mealser

ETH Zurich
Fall 2022
31 October 2022

SAFARI

ETH zürich

Agenda for Today

- What is Genome Analysis?
- What is Intelligent Genome Analysis?
- How we Analyze Genome?

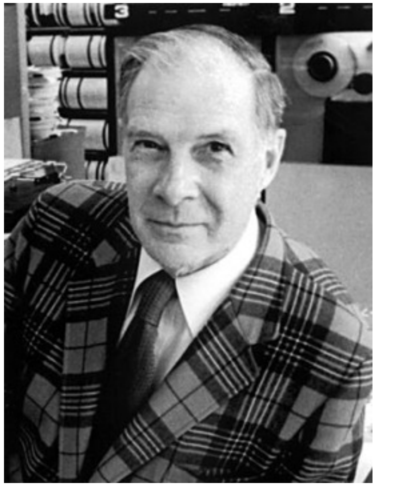
Agenda for Today

- **What is Genome Analysis?**
- What is Intelligent Genome Analysis?
- How we Analyze Genome?

What is Data Analysis?

"The purpose of computing is [to gain] insight, not numbers"

Richard Hamming



What is Genome Analysis?

https://onlinelearning.hms.harvard.edu/hmx/courses/genetic-testing/
https://www.nature.com/subjects/genomic-analysis

What is Genome Analysis?

nature research

Search Login

nature > subjects > genomic analysis

Genomic analysis

Atom RSS Feed

Genomic analysis is the identification, measurement or comparison of genomic features such as DNA sequence, structural variation, gene expression, or regulatory and functional element annotation at a genomic scale. Methods for genomic analysis typically require high-throughput sequencing or microarray hybridization and bioinformatics.

https://onlinelearning.hms.harvard.edu/hmx/courses/genetic-testing/
https://www.nature.com/subjects/genomic-analysis

DNA Testing

Health + Ancestry Service

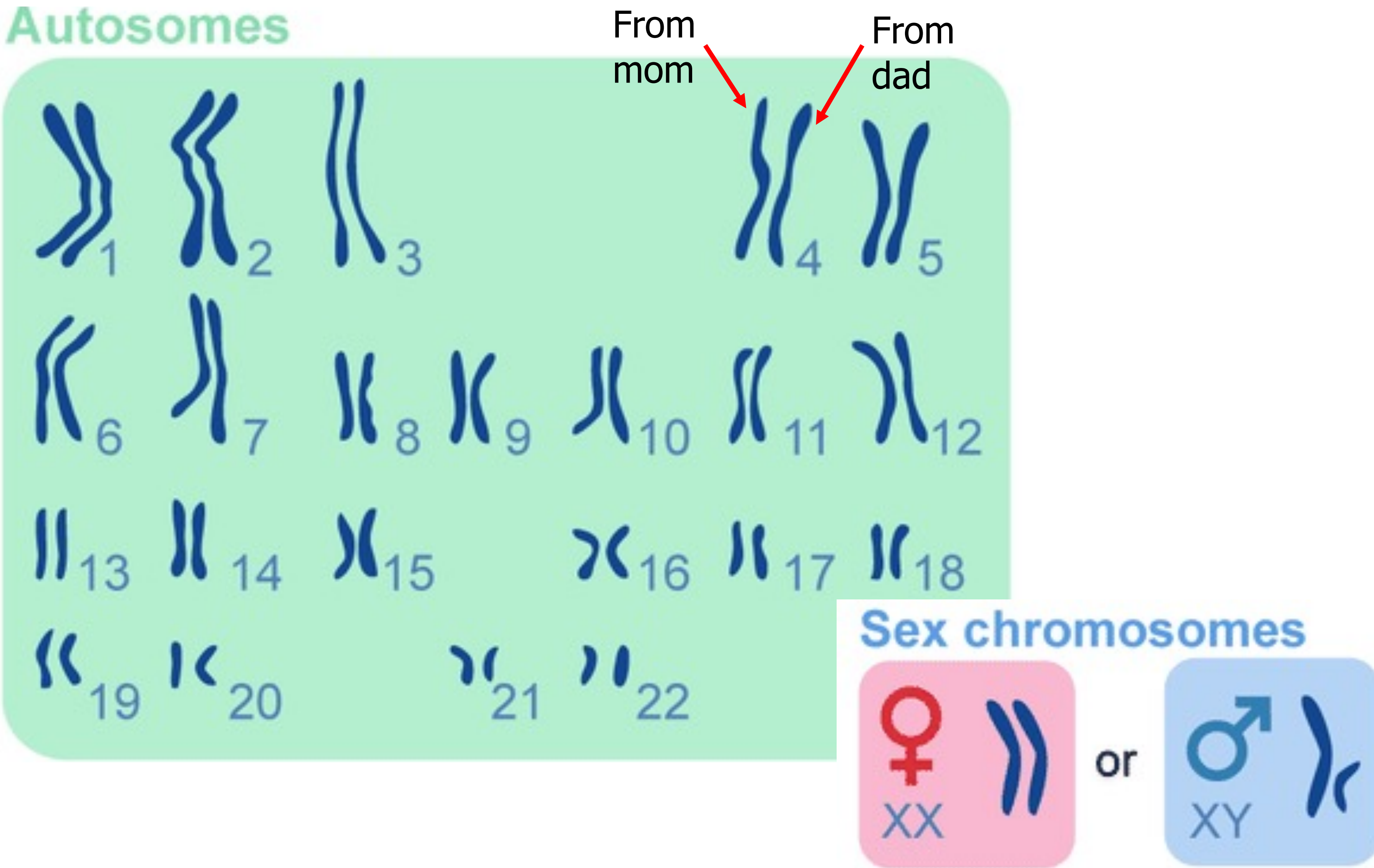
$199

- Includes everything in Ancestry + Traits Service

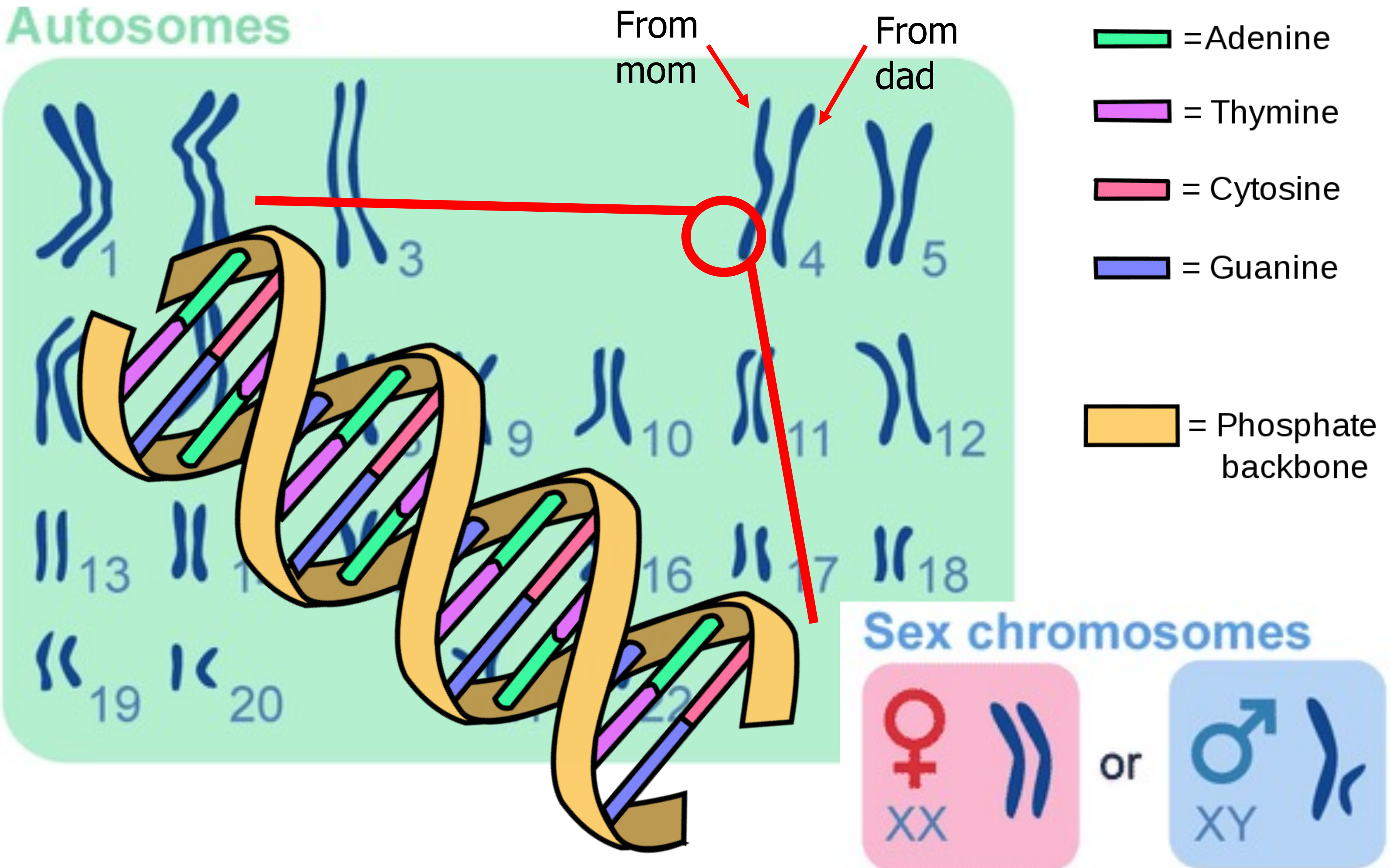
PLUS

- 10+ Health Predisposition reports*
- 5+ Wellness reports
- 40+ Carrier Status reports*

Human Chromosomes (23 Pairs)



Human Chromosomes (23 Pairs)



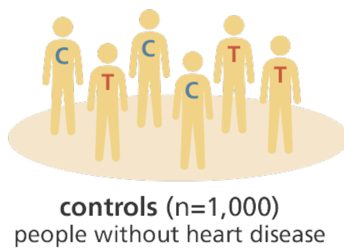
Finding SNPs Associated with Complex Trait

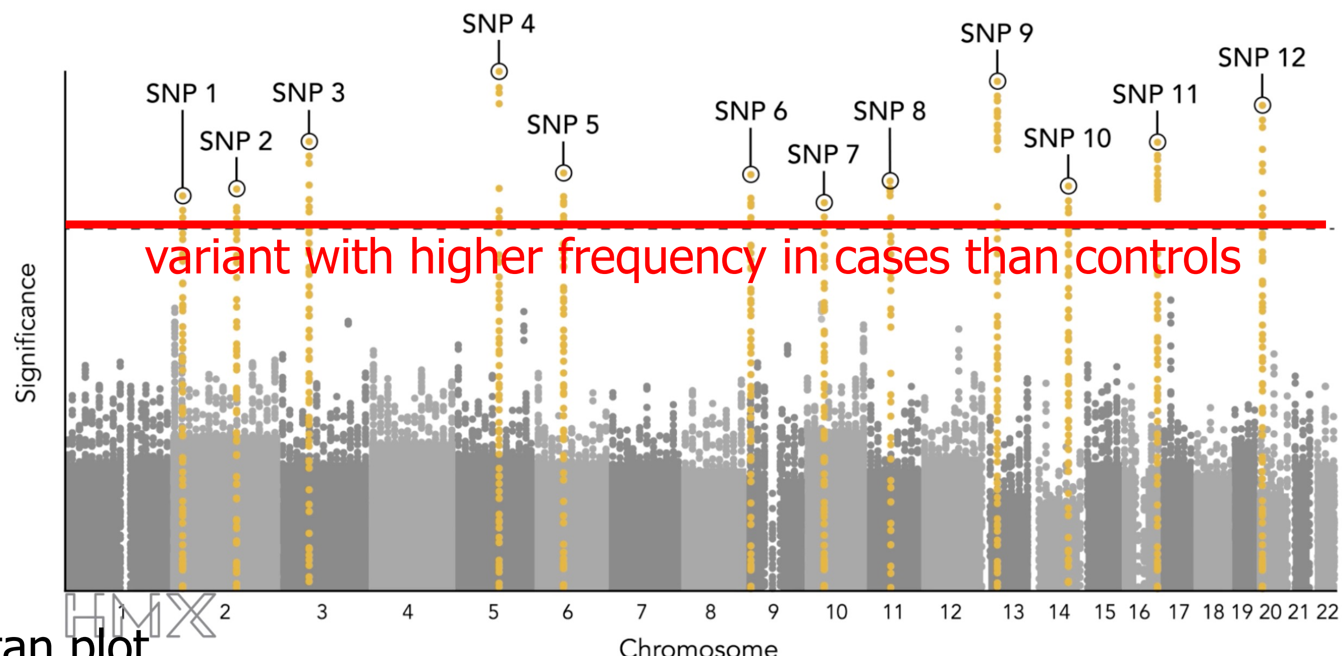
	SNP1 SNP2	Blood Pressure
Individual #1	**...ACATGCCGACATTTCATAGGCC...**	180
Individual #2	**...ACATGCCGACATTTCATAAGCC...**	175
Individual #3	**...ACATGCCGACATTTCATAGGCC...**	170
Individual #4	**...ACATGCCGACATTTCATAAGCC...**	165
Individual #5	**...ACATGCCGACATTTCATAGGCC...**	160
Individual #6	**...ACATGCCGACATTTCATAGGCC...**	145
Individual #7	**...ACATGCCGACATTTCATAAGCC...**	140
Individual #8	**...ACATGCCGACATTTCATAAGCC...**	130
Individual #9	**...ACATGTCGACATTTCATAGGCC...**	120
Individual #10	**...ACATGTCGACATTTCATAAGCC...**	120
Individual #11	**...ACATGTCGACATTTCATAGGCC...**	115
Individual #12	**...ACATGTCGACATTTCATAAGCC...**	110
Individual #13	**...ACATGTCGACATTTCATAGGCC...**	110
Individual #14	**...ACATGTCGACATTTCATAAGCC...**	110
Individual #15	**...ACATGTCGACATTTCATAGGCC...**	105
Individual #16	**...ACATGTCGACATTTCATAAGCC...**	100

SNP: single nucleotide polymorphism

Genome-Wide Association Study (GWAS)

- Detecting genetic variants associated with phenotypes using two groups of people.





Manhattan plot

Similar Association Studies

PERSPECTIVE
https://doi.org/10.1038/s41588-019-0385-z
nature genetics

Opportunities and challenges for transcriptome-wide association studies

Michael Wainberg[1], Nasa Sinnott-Armstrong[2], Nicholas Mancuso[3], Alvaro N. Barbeira[4], David A. Knowles[5,6], David Golan[2], Raili Ermel[7], Arno Ruusalepp[7,8], Thomas Quertermous[9], Ke Hao[10], Johan L. M. Björkegren[8,10,11,12*], Hae Kyung Im[4*], Bogdan Pasaniuc[3,13,14*], Manuel A. Rivas[15*] and Anshul Kundaje[1,2*]

Transcriptome-wide association studies (TWAS) integrate genome-wide association studies (GWAS) and gene expression datasets to identify gene-trait associations. In this Perspective, we explore properties of TWAS as a potential approach to prioritize causal genes at GWAS loci, by using simulations and case studies of literature-curated candidate causal genes for schizophrenia, low-density-lipoprotein cholesterol and Crohn's disease. We explore risk loci where TWAS accurately prioritizes the likely causal gene as well as loci where TWAS prioritizes multiple genes, some likely to be non-causal, owing to sharing of expression quantitative trait loci (eQTL). TWAS is especially prone to spurious prioritization with expression data from non-trait-related tissues or cell types, owing to substantial cross-cell-type variation in expression levels and eQTL strengths. Nonetheless, TWAS prioritizes candidate causal genes more accurately than simple baselines. We suggest best practices for causal-gene prioritization with TWAS and discuss future opportunities for improvement. Our results showcase the strengths and limitations of using eQTL datasets to determine causal genes at GWAS loci.

Wainberg+, "Opportunities and challenges for transcriptome-wide association studies", *Nature genetics,* 2019.

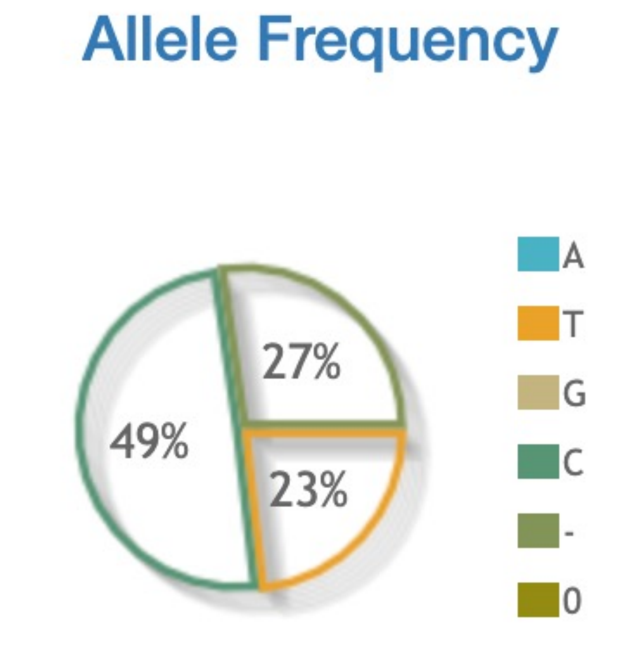
SNPs and Personalized Medicine

openSNP Search

SNP rs12979860

Basic Information

Name	rs12979860
Chromosome	19
Position	39248147
Weight of evidence	926



Links to SNPedia

Title	Summary
rs12979860 T/T	~20-25% of such hepatitis c patients respond to treatment
rs12979860 C/C	~80% of such hepatitis c patients respond to treatment
rs12979860 C/T	~20-40% of such hepatitis c patients respond to treatment

Personalized Medicine for Critically Ill Infants

- rWGS can be performed in 2-day (costly) or 5-day time to interpretation.
- Diagnostic rWGS for infants
  - Avoids morbidity
  - Reduces hospital stay length by 6%-69%
  - Reduces inpatient cost by $800,000-$2,000,000.

Article | Open Access | Published: 04 April 2018

Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization

Lauge Farnaes, Amber Hildreth, Nathaly M. Sweeney, Michelle M. Clark, S
Chowdhury, Shareef Nahas, Julie A. Cakici, Wendy Benson, Robert H. Kap
Richard Kronick, Matthew N. Bainbridge, Jennifer Friedman, Jeffrey J. Go
Ding, Narayanan Veeraraghavan, David Dimmock & Stephen F. Kingsmore

npj Genomic Medicine **3**, Article number: 10 (2018) | Cite this article

Article | Open Access | Published: 05 May 2020

Clinical utility of 24-h rapid trio-exome sequencing for critically ill infants

Huijun Wang, Yanyan Qian, Yulan Lu, Qian Qin, Guoping Lu, Guoqiang Cheng, Ping Zhang, Lin Yang, Bingbing Wu & Wenhao Zhou

npj Genomic Medicine **5**, Article number: 20 (2020) | Cite this article

Farnaes+, "Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization", NPJ Genomic Medicine, 2018

Personalized Medicine in UK

"From 2019, all seriously ill children in UK will be offered whole genome sequencing as part of their care"

NHS
National Institute for Health Research

Much Larger Structural Variations!

AUTISM
Weiss, *N Eng J Med* 2008
Deletion of 593 kb

SCHIZOPHRENIA
McCarthy, *Nat Genet* 2009
Duplication of 593 kb

OBESITY
Walters, *Nature* 2010
Deletion of 593 kb

UNDERWEIGHT
Jacquemont, *Nature* 2011
Duplication of 593 kb

Deletion in the short arm of chromosome 16 (16p11.2)

Duplication in the short arm of chromosome 16 (16p11.2)

CNV: copy number variation

Recommended Reading

nature reviews genetics

Explore our content ⌄ Journal information ⌄

nature > nature reviews genetics > review articles > article

Review Article | Published: 15 November 2019

Structural variation in the sequencing era

Steve S. Ho, Alexander E. Urban & Ryan E. Mills

Nature Reviews Genetics **21**, 171–189(2020) | Cite this article

15k Accesses | **16** Citations | **309** Altmetric | Metrics

Ho+, "Structural variation in the sequencing era", Nature Reviews Genetics, 2020

Agenda for Today

- What is Genome Analysis?
- **What is Intelligent Genome Analysis?**
- How we Analyze Genome?

What is Intelligent Genome Analysis?

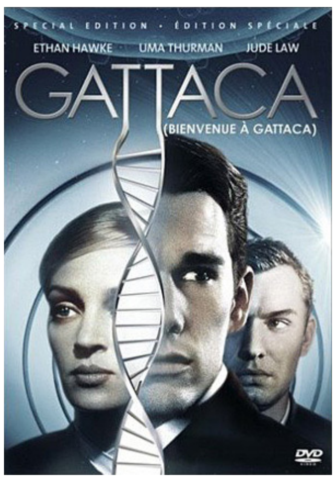
- Fast genome analysis — **Bandwidth**
  - *Real-time analysis*
- Using intelligent architectures — **Energy-efficiency & Latency**
  - *Specialized HW with less data movement*
- DNA is a valuable asset — **Privacy**
  - *Controlled-access analysis*
- Population-scale genome analysis — **Scalability**
  - *Sequence anywhere at large scale!*
- Avoiding erroneous analysis — **Accuracy**
  - *E.g., your father is not your father*

Does intelligent genome analysis really matter?

Fast Genome Analysis?

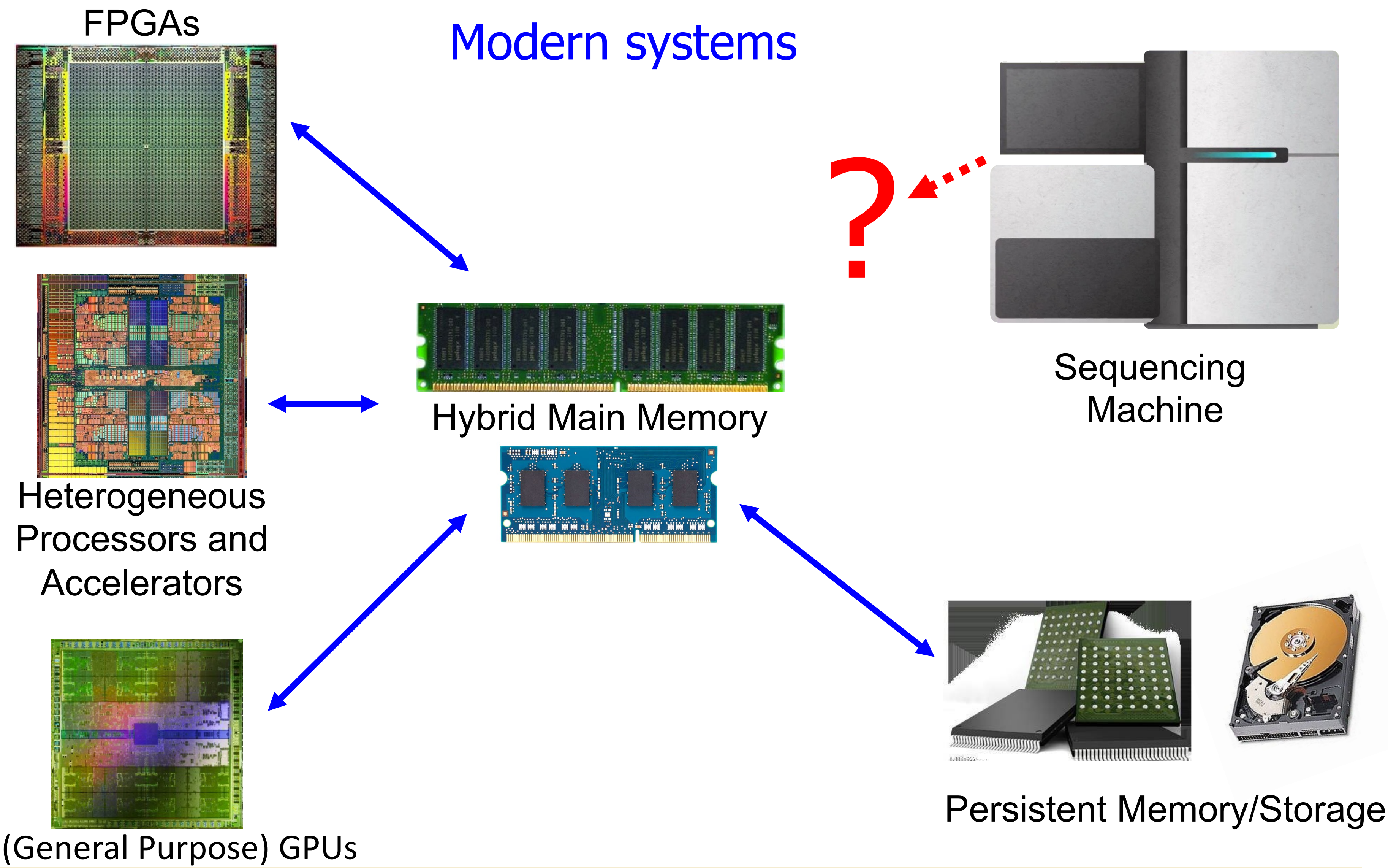
- **Fast** genome analysis in mere seconds using limited computational resources (i.e., personal computer or small hardware).

1997

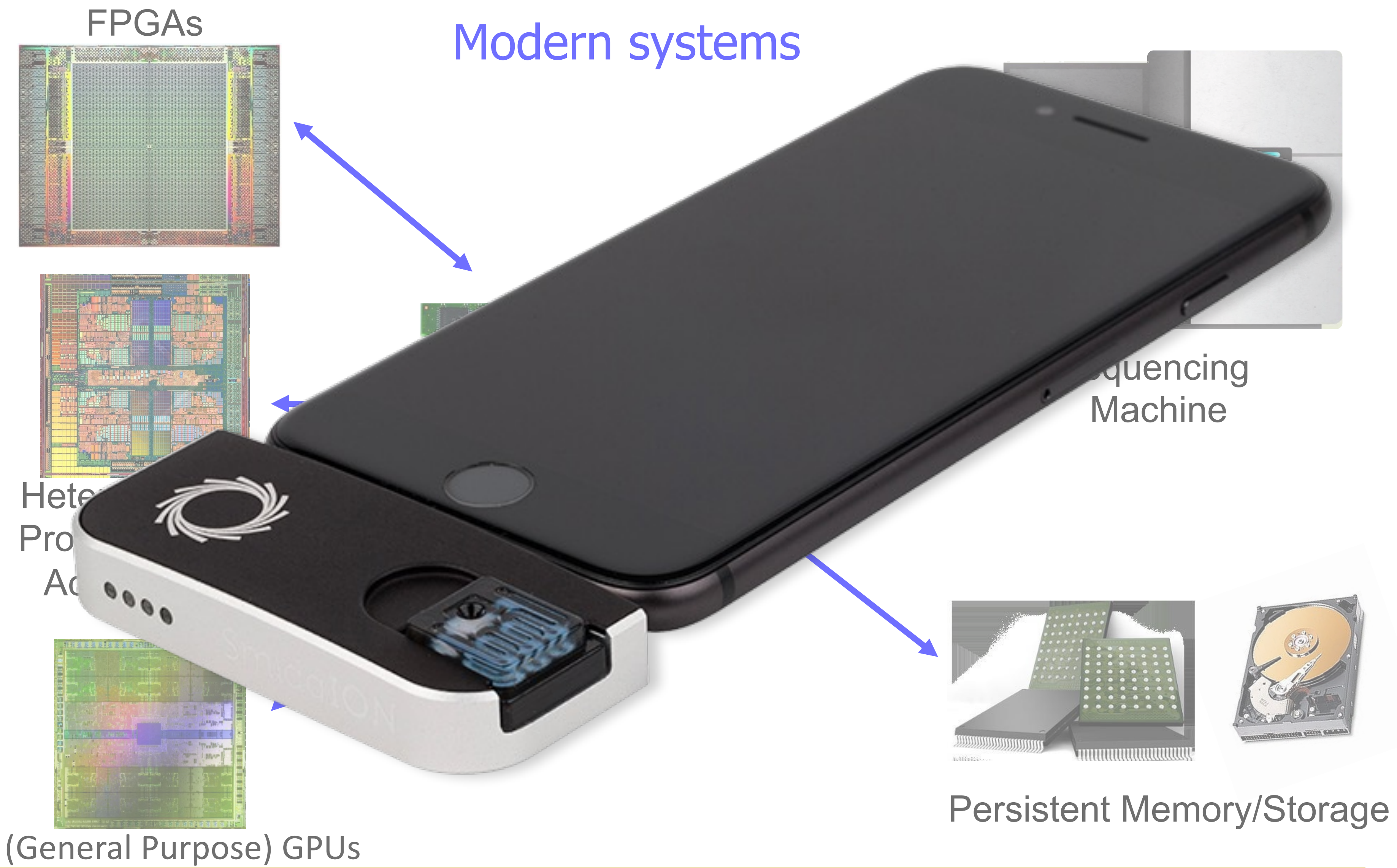


2015

Intelligent Architecture?



Intelligent Architecture?



https://nanoporetech.com/products/smidgion

Privacy-Preserving Genome Analysis?

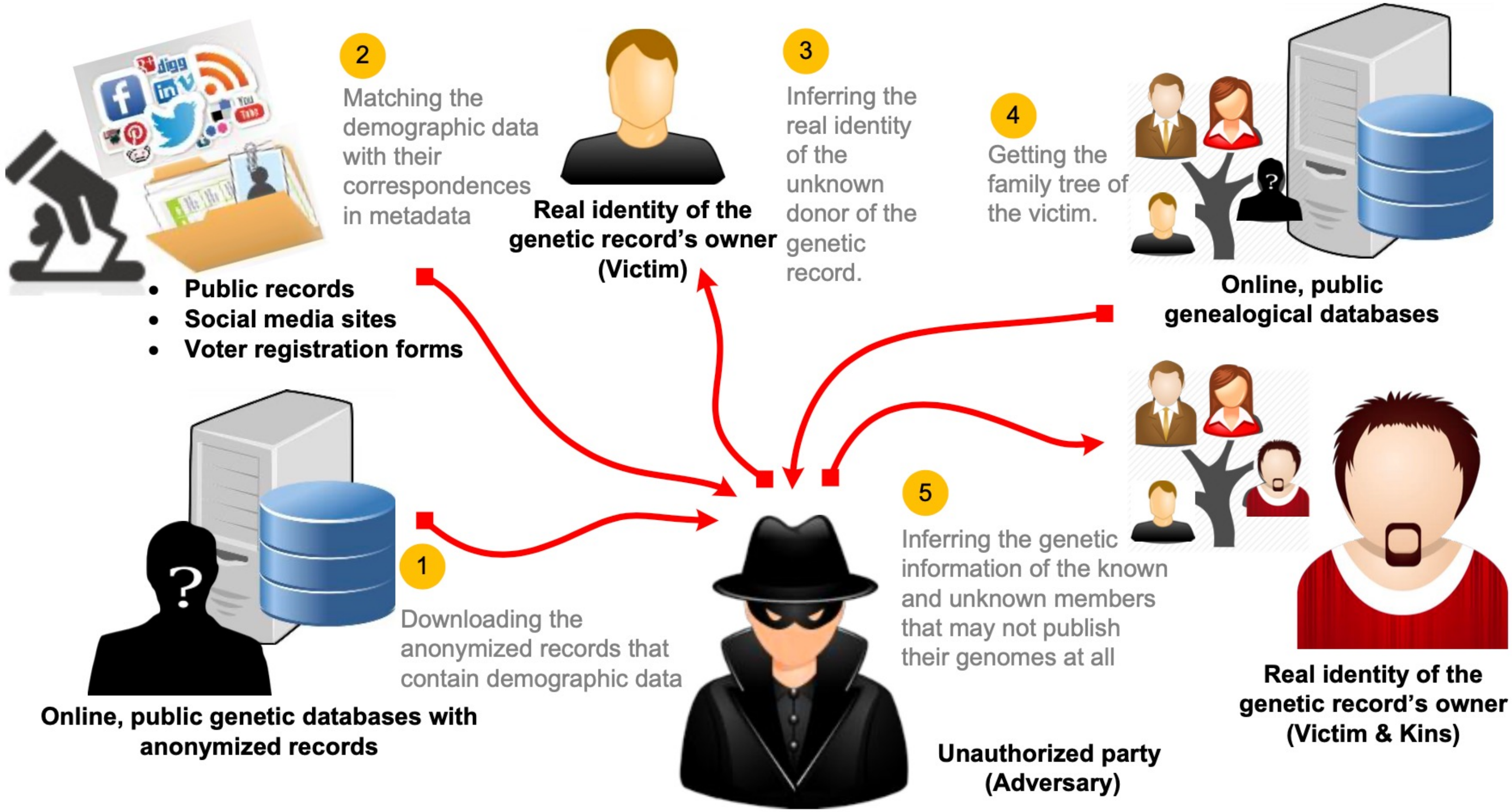


Fig. 5. A completion attack.

Alser+, "[Can you really anonymize the donors of genomic data in today's digital world?]" *10th International Workshop on Data Privacy Management (DPM)*, 2015.

Can you Really Anonymize the Donors?

(Position Paper) Can You Really Anonymize the Donors of Genomic Data in Today's Digital World?

Mohammed Alser, Nour Almadhoun, Azita Nouri, Can Alkan, and Erman Ayday

Computer Engineering Department, Bilkent University, 06800 Bilkent, Ankara, Turkey

Abstract. The rapid progress in genome sequencing technologies leads to availability of high amounts of genomic data. Accelerating the pace of biomedical breakthroughs and discoveries necessitates not only collecting millions of genetic samples but also granting open access to genetic databases. However, one growing concern is the ability to protect the privacy of sensitive information and its owner. In this work, we survey a wide spectrum of cross-layer privacy breaching strategies to human genomic data (using both public genomic databases and other public non-genomic data). We outline the principles and outcomes of each technique, and assess its technological complexity and maturation. We then review potential privacy-preserving countermeasure mechanisms for each threat.

Keywords: Genomics, Privacy, Bioinformatics

DPM 2015
Vienna, Austria
September 21-22, 2015

Alser+, "**Can you really anonymize the donors of genomic data in today's digital world?**" *10th International Workshop on Data Privacy Management (DPM)*, 2015.

Privacy-Preserving DNA Test

Our DNA Test, Reports, and Technology

- **Whole Genome Sequencing.** Decode 100% of your DNA with Whole Genome Sequencing and fully unlock your genetic blueprints.
- **Privacy First DNA Testing.** Begin your journey of discovery without risking the privacy of your most personal information.
- **Nebula Research Library.** Receive new reports every week that are based on the latest scientific discoveries.
- **Genome Exploration Tools.** Use powerful, browser-based genome exploration tools to answer any questions about your DNA.
- **Deep Genetic Ancestry.** Discover more about your ancestry with full Y chromosome and mitochondrial DNA sequencing and analysis.
- **Genomic Big Data Access.** Download your FASTQ, BAM, and VCF files and dive deeper into your Whole Genome Sequencing data.
- **Ready for Diagnostics.** Our Whole Genome Sequencing data is of the highest quality and can be used by physicians and genetic counselors.

30x Whole Genome Sequencing DNA Test — **$299** Normally ~~$1000~~ Save 70%!

A genetic test that decodes 100% of your DNA with very high accuracy. 30x Whole Genome Sequencing offers the best value for money and is the best choice for most people.

100x Whole Genome Sequencing DNA Test — **$999** Normally ~~$3500~~ Save 70%!

A genetic test that decodes 100% of your DNA with extremely high accuracy. 100x Whole Genome Sequencing is recommended for the discovery of rare genetic mutations.

Get Sequenced

Rapid Surveillance of Disease Outbreaks?

Figure 1: Deployment of the portable genome surveillance system in Guinea.

Quick+, "[Real-time, portable genome sequencing for Ebola surveillance]", *Nature*, 2016

Scalable SARS-CoV-2 Testing

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

HOME | ABOU

Search

Comments (1)

Swab-Seq: A high-throughput platform for massively scaled up SARS-CoV-2 testing

Joshua S. Bloom, Eric M. Jones, Molly Gasperini, Nathan B. Lubock, Laila Sathe, Chetan Munugala, A. Sina Booeshaghi, Oliver F. Brandenberg, Longhua Guo, James Boocock, Scott W. Simpkins, Isabella Lin, Nathan LaPierre, Duke Hong, Yi Zhang, Gabriel Oland, Bianca Judy Choe, Sukantha Chandrasekaran, Evann E. Hilt, Manish J. Butte, Robert Damoiseaux, Aaron R. Cooper, Yi Yin, Lior Pachter, Omai B. Garner, Jonathan Flint, Eleazar Eskin, Chongyuan Luo, Sriram Kosuri, Leonid Kruglyak, Valerie A. Arboleda

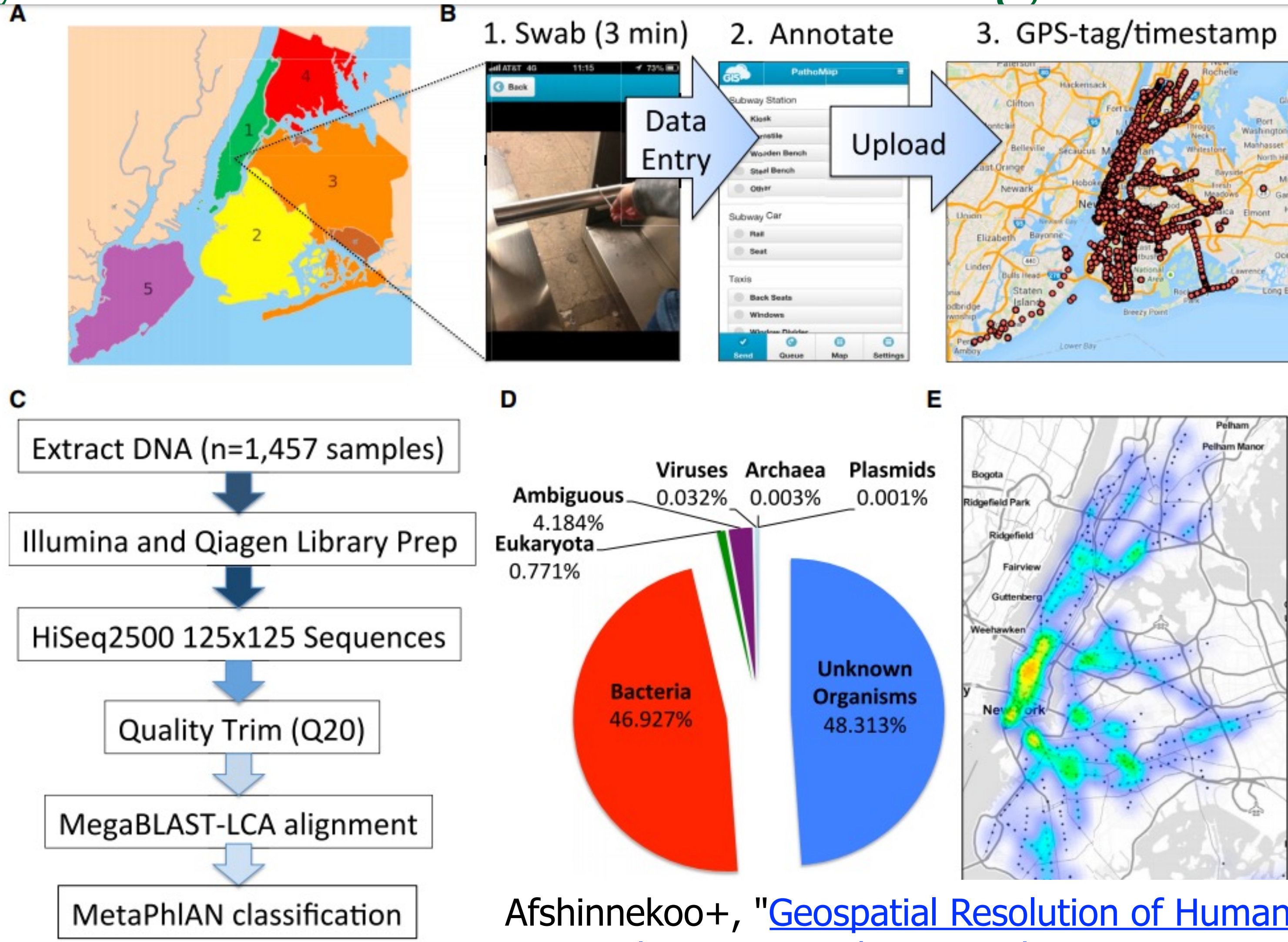
doi: https://doi.org/10.1101/2020.08.04.20167874

Bloom+, "Swab-Seq: A high-throughput platform for massively scaled up SARS-CoV-2 testing", *medRxiv*, 2020

Population-Scale Microbiome Profiling

SAFARI https://blog.wego.com/7-crowded-places-and-events-that-you-will-love/

City-Scale Microbiome Profiling



Afshinnekoo+, "[Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics]", Cell Systems, 2015

Figure 1. The Metagenome of New York City

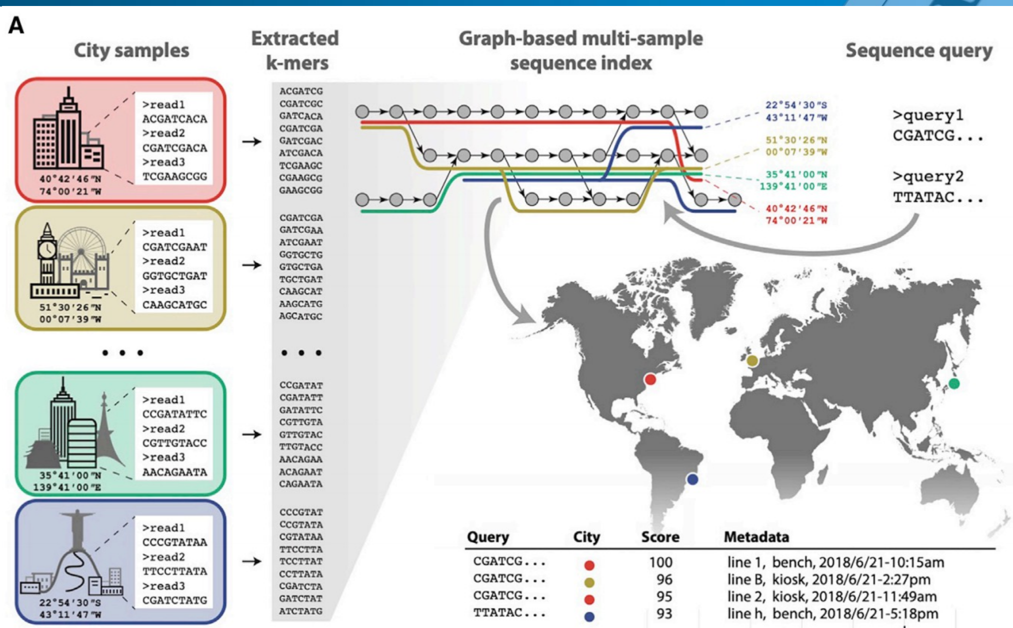
(A) The five boroughs of NYC include (1) Manhattan (green)

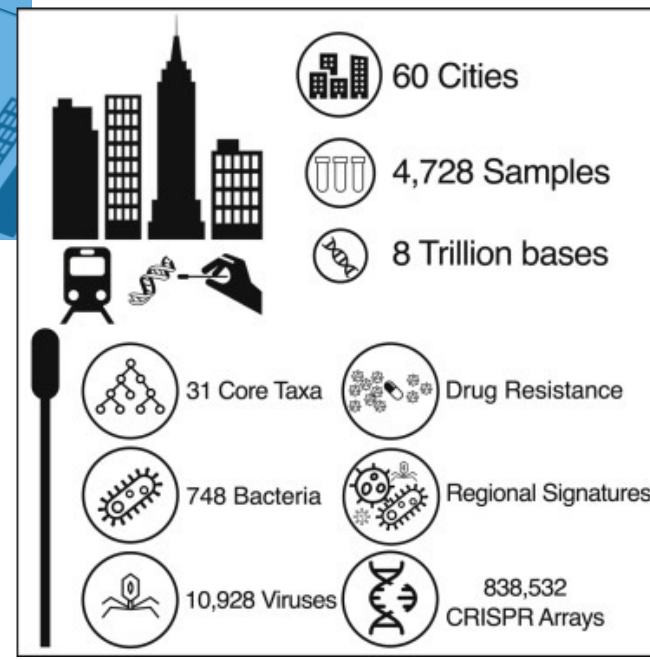
(B) The collection from the 466 subway stations of NYC across the 24 subway lines involved three main steps: (1) collection with Copan Elution swabs, (2) data entry into the database, and (3) uploading of the data. An image is shown of the current collection database, taken from http://pathomap.giscloud.com.

(C) Workflow for sample DNA extraction, library preparation, sequencing, quality trimming of the FASTQ files, and alignment with MegaBLAST and MetaPhlAn to discern taxa present

SAI

Population-Scale Microbiome Profiling





Danko+, "A global metagenomic map of urban microbiomes and antimicrobial resistance", Cell, 2021

Plague in New York Subway System?

Plague (Yersinia Pestis)

Harvard Health Publishing
HARVARD MEDICAL SCHOOL
Trusted advice for a healthier life

What Is It?

Published: December, 2018

Plague is caused by Yersinia pestis bacteria. It can be a life-threatening infection if not treated promptly. Plague has caused several major epidemics in Europe and Asia over the last 2,000 years. Plague has most famously been called "the Black Death" because it can cause skin sores that form black scabs. A plague epidemic in the 14th century killed more than one-third of the population of Europe within a few years. In some cities, up to 75% of the population died within days, with fever and swollen skin sores.

Plague in New York Subway System?

Plague (Yersi

What Is It?

Published: December, 2018

Plague is caused by Yersinia
treated promptly. Plague ha
last 2,000 years. Plague has
cause skin sores that form b
than one-third of the popul
the population died within

The New York Times

Bubonic Plague in the Subway System? Don't Worry About It

In October, riders were not deterred after reports that an Ebola-infected man had ridden the subway just before he fell ill. Robert Stolarik for The New York Times

https://www.nytimes.com/2015/02/07/nyregion/bubonic-plague-in-the-subway-system-dont-worry-about-it.html

The findings of Yersinia Pestis in the subway received wide coverage in the lay press, causing some alarm among New York residents

Failure of Bioinformatics

data. Rob Knight, a professor in the department of pediatrics at the University of California, San Diego, calls this type of error “a failure of bioinformatics,” in that Mason had assumed the gene fragments were unique to the pathogens, when in fact they can also be detected in other

Living in a microbial world
Charles Schmidt
Nature Biotechnology, **volume 35**, pages401–403 (2017)
https://www.nature.com/articles/nbt.3868

There is a critical need for **fast** and **accurate** genome analysis.

Achieving Intelligent Genome Analysis?

How and where to enable

fast, accurate, cheap,

privacy-preserving, and exabyte scale

analysis of genomic data?

Agenda for Today

- What is Genome Analysis?

- What is Intelligent Genome Analysis?

- **How we Analyze Genome?**

Genome Analysis

NO machine can read the ***entire*** content of a genome

>CCTCCTCAGTGCCACCCAGCCCACTGGCAGCTCCCAAACAGGCTCTTATTAAAACACCCTGTTCCCTGCCCCTTGGAGTGAGGTGTCAAG
GACCTAAACTAAAAAAAAAAAAAAGAAAAAGAAAAGAAAAAGAATTTAAAATTTAAGTAATTCTTTGAAAAAAACTAATTTCTAAGCTTCTT
CATGTCAAGGACCTAATGTGCTAAACAGCACTTTTTTGACCATTATTTTGGATCTGAAAGAAATCAAGAATAAATGAAGGACTTGATACATTG
GAAGAGGAGAGTCAAGGACCTACAGAAAAAAAAAAAAAAAGAAAAAGAAAAGAAAAAGA**A**TTTAAAATTTAAGTAATTCTTTGAAAAAA
ACTAATTTCTAAGCTTCTT**C**ATGTCAAGGACCTAATGTCTGTGTTGCAGGTCTTCTTGCATTTCCCTGTCAAAAGAAAAAGAATTTAAAATTT
AAGTAATTCTTTGAAAAAAACTAATTTCTAAGCTTCTTCATGTCAAGGACCTAATGTCAGGCCAAGAGTTGCAAAAAAAAAAAAAAGAAAAA
GAAAAGAAAAAGAATTTAAAATTTA**A**GTAATTCTTTGAAAAAAACTAATTTCTAAGCTTCTTCATGTCAAGGACCTAATGTAGCCAGAATGG
TTGTGGGATGGGAGCCTCTGTGGACCGACCAGGTAGCTCTCTTTTCCACACTGTAGTCTCAAAGCTTCTTCATGTGGTTTCTCTGAGTGAAA
AAAAAAAAAAGAAAAAGAAAAGAAAAAGAATTTAAAATTTAAGTAATTCTTTGAAAAAAACTAATTTCTAAGCTT**T**TTCATGTCAAGGACC
TAATGTAGCTATACTGAACGTTATCTAGGGGAAAGATTGAAGGGGAGCTCTAAGGTCAACACACCACCACTTCCCAGAAAGCTTCTTCA……

Genome Analysis

NO machine can read the ***entire*** content of a genome

Why?!

>CCT ... CAAG
GACC ... TCTT
CATGT ... CATTG
GAAG ... AAAA
ACTA ... ATTT
AAGT ... AAAA
GAAA ... ATGG
TTGTC ... GAAA
AAAAAAAAAGAAAAAGAAAAGAAAAAGAATTTAAAATTTAAGTAATTCTTTGAAAAAAACTAATTTCTAAGCTTTTTCATGTCAAGGACC
TAATGTAGCTATACTGAACGTTATCTAGGGGAAAGATTGAAGGGGAGCTCTAAGGTCAACACACCACCACTTCCCAGAAAGCTTCTTCA......

Suggested Readings

nature methods

Explore content ˅ About the journal ˅ Publish with us ˅

Published: November 2009

Next-generation sequencing library preparation: simultaneous fragmentation and tagging using *in vitro* transposition

Fraz Syed, Haiying Grunenwald & Nicholas Caruccio

Nature Methods **6**, i–ii (2009) | Cite this article

https://www.nature.com/articles/nmeth.f.272

Suggested Readings

nature biotechnology

Explore content ⌄ **About the journal** ⌄ **Publish with us** ⌄

nature › nature biotechnology › review articles › article

Published: 09 October 2008

Next-generation DNA sequencing

Jay Shendure & Hanlee Ji

Nature Biotechnology **26**, 1135–1145 (2008) | Cite this article

https://www.nature.com/articles/nbt1486

Genome Sequencer is a Chopper

GATK

CCCCCCTATATATACGTACTAGTACGT
ACGACTTTAGTACGTACGT
TATATATACGTACTAGTACGT
ACGTACG CCCCTACGTA
TATATATACGTACTAGTACGT
ACGACTTTAGTACGTACGT
TATATATACGTACTAAAGTACGT
TATATATACGTACTAGTACGT
ACG TTTTTAAAACGTA
TATATATACGTACTAGTACGT
ACGAC GGGGAGTACGTACGT

$1\text{x}10^{12}$ bases*

44 hours*

<1000 $

* NovaSeq 6000

High-Throughput Sequencers

Illumina MiSeq

Pacific Biosciences Sequel II

Oxford Nanopore PromethION

Oxford Nanopore MinION

Illumina NovaSeq 6000

Pacific Biosciences RS II

Oxford Nanopore SmidgION

... and more! All produce data with different properties.

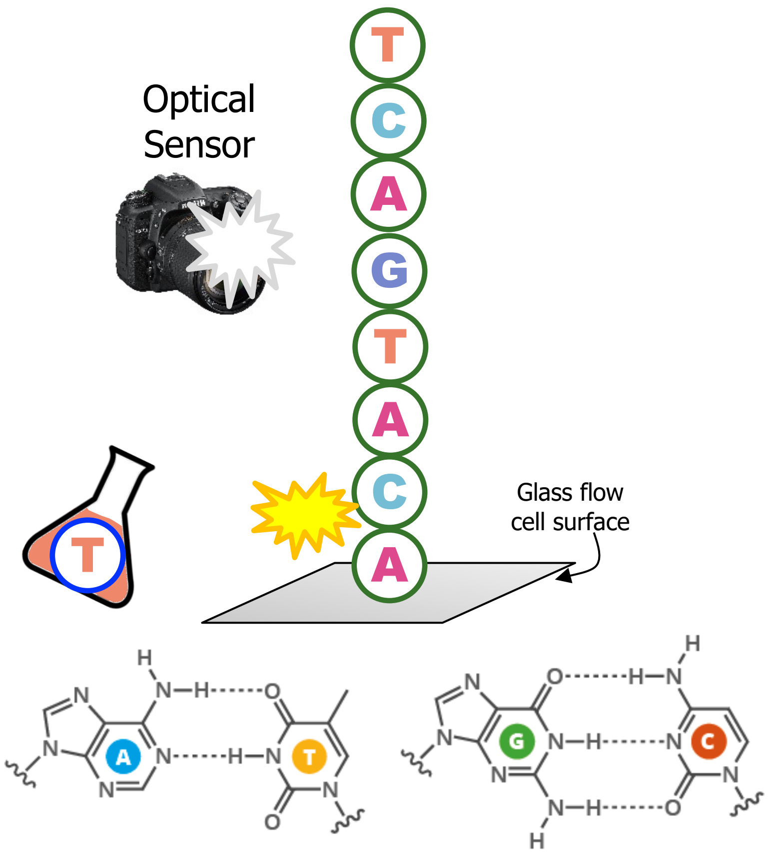
Oxford Nanopore Sequencers

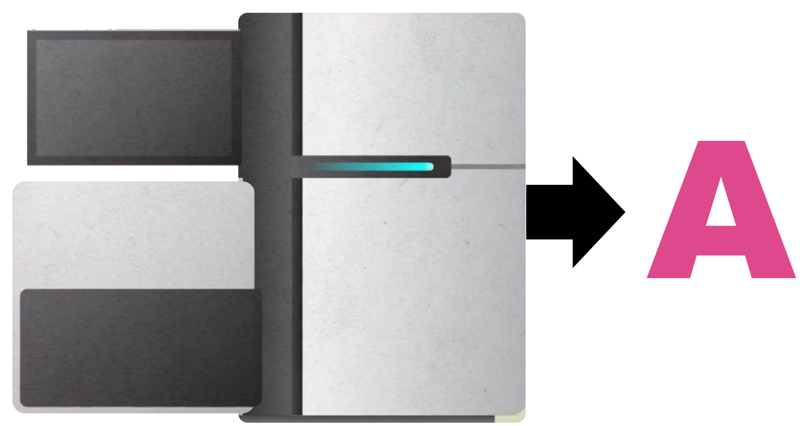
	MinION Mk1B	MinION Mk1C	GridION Mk1	PromethION 24	PromethION 48
Read length	> 2Mb	> 2Mb	> 2Mb	> 2Mb	> 2Mb
Yield per flow cell	50 Gb	50 Gb	50 Gb	220 Gb	220 Gb
Number of flow cells per device	1	1	5	24	48
Yield per device	<50 Gb	<50 Gb	<250 Gb	<5.2 Tb	<10.5 Tb
Starting price	$1,000	$4,990	$49,995	$195,455	$327,455

Illumina Sequencers

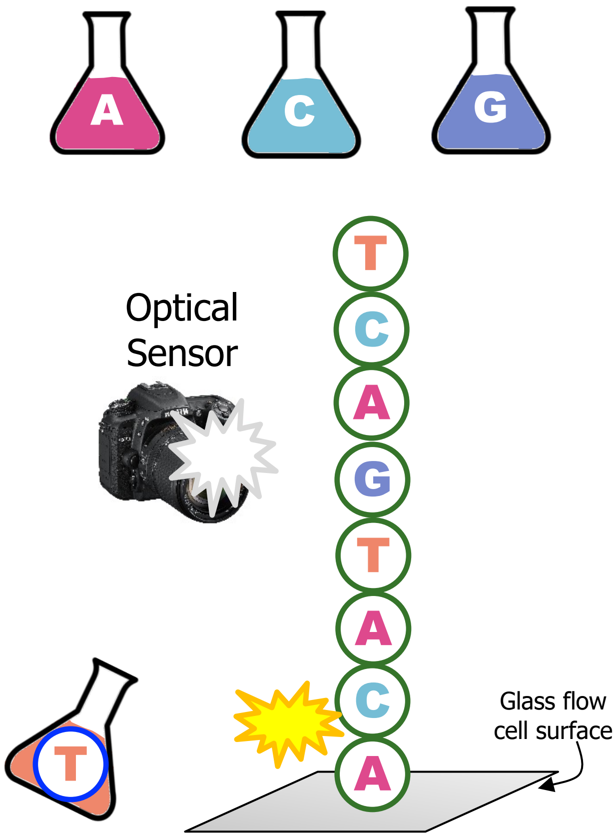
Run time	9.5–19 hrs	4–24 hrs	4–55 hrs	12–30 hrs	24-48 hrs	13-44 hrs
Max. reads per run	4 million	25 million	25 million	400 million	1 billion	20 billion
Max. read length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 x 250
Max. output	1.2 Gb	7.5 Gb	15 Gb	120 Gb	300 Gb	6000 Gb
Estimated price	$19,900	$49,500	$128,000	$275,000	$335,000	$985,000

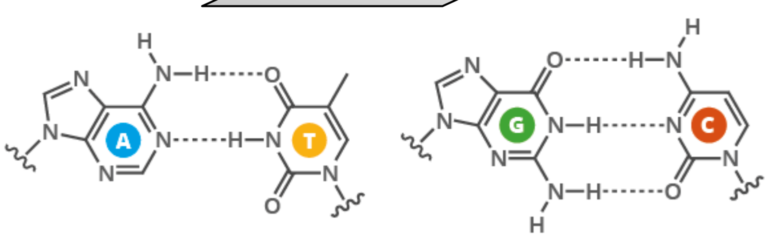
How Does Illumina Machine Work?

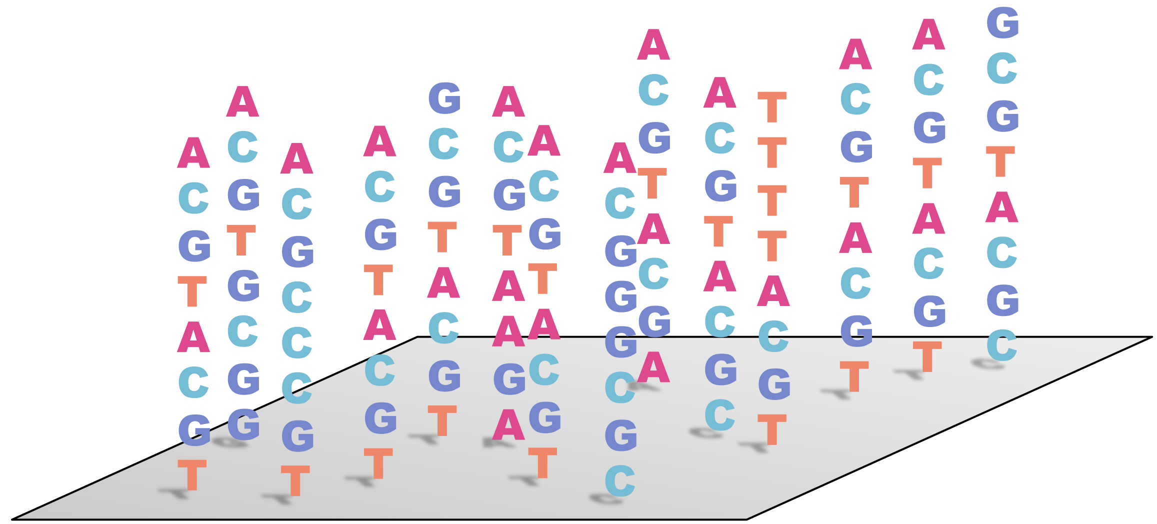


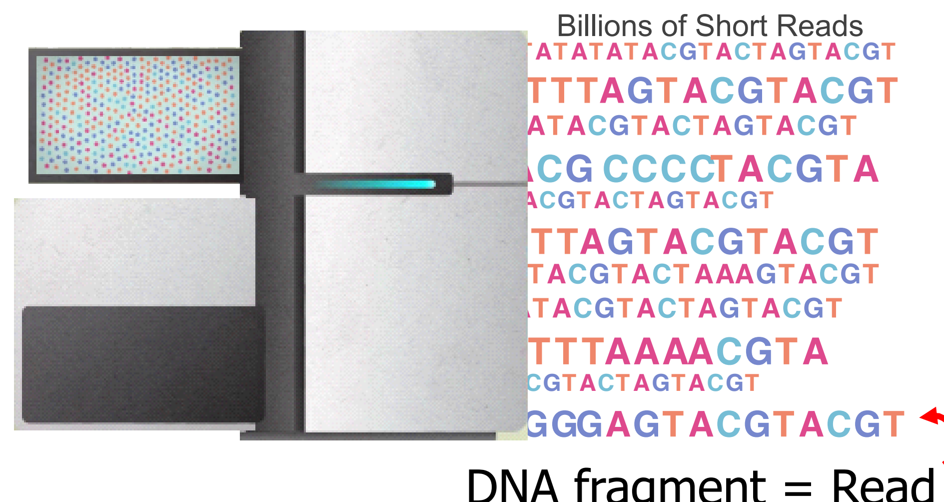


How Does Illumina Machine Work?

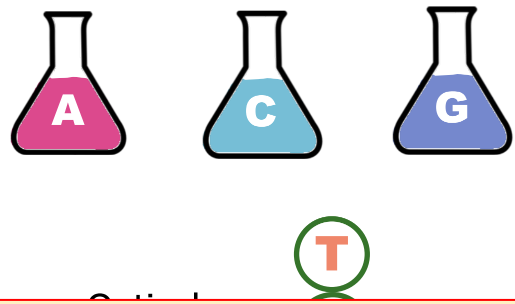






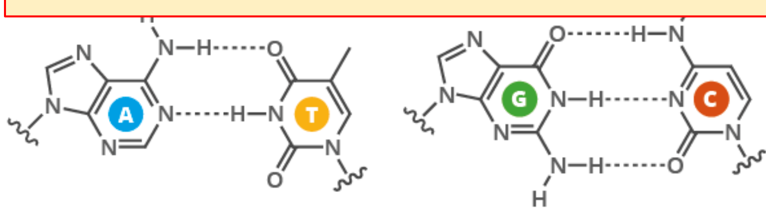


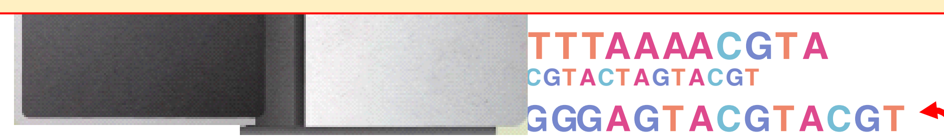
How Does Illumina Machine Work?



Check Illumina virtual tour:

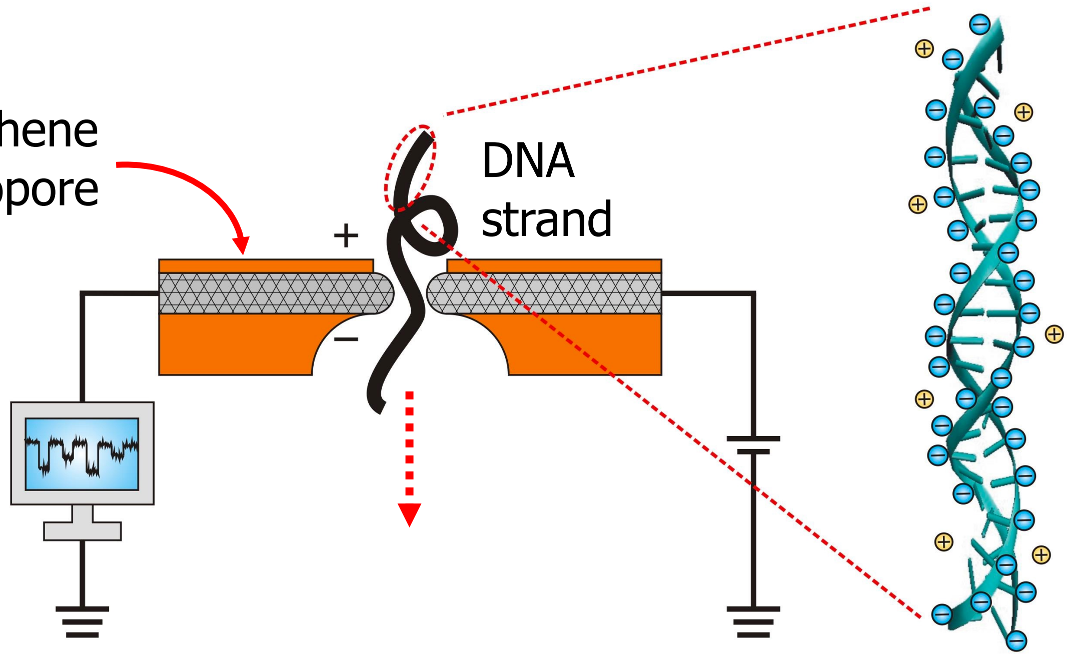
https://emea.illumina.com/systems/sequencing-platforms/iseq/tour.html





DNA fragment = Read

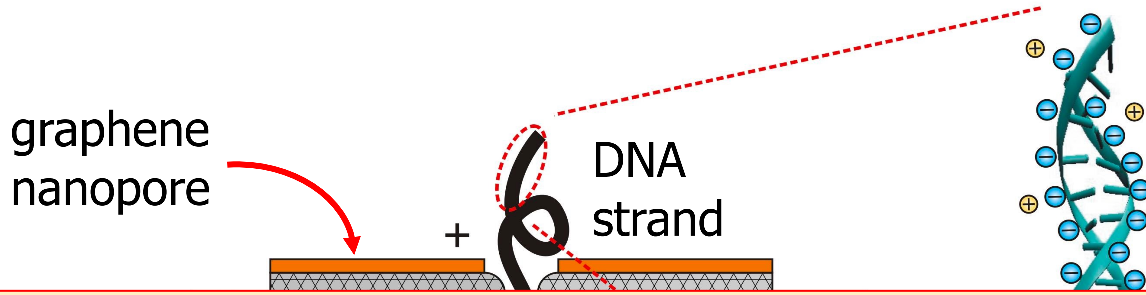
How Does Nanopore Machine Work?



- **Nanopore** is a nano-scale hole (<20nm).
- In nanopore sequencers, an **ionic current** passes through the nanopores
- When the DNA strand passes through the nanopore, the sequencer measures the the **change in current**
- This change is used to identify the bases in the strand with the help of **different electrochemical structures** of the different bases

Figure is adapted from: https://phys.org/news/2013-12-gene-sequencing-future.html

How Does Nanopore Machine Work?



Check Nanopore virtual tour:

https://nanoporetech.com/resource-centre/minion-video

measures the the **change in current**

- This change is used to identify the bases in the strand with the help of **different electrochemical structures** of the different bases

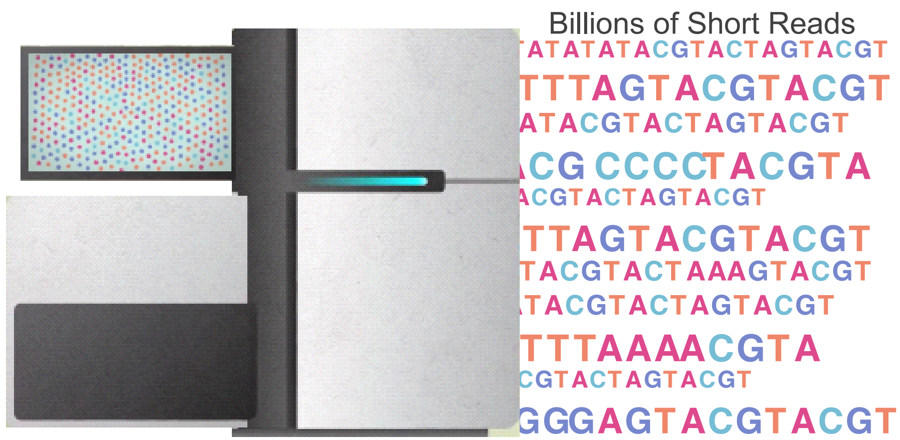
Figure is adapted from: https://phys.org/news/2013-12-gene-sequencing-future.html

Common Disadvantages!

Regardless the sequencing machine,

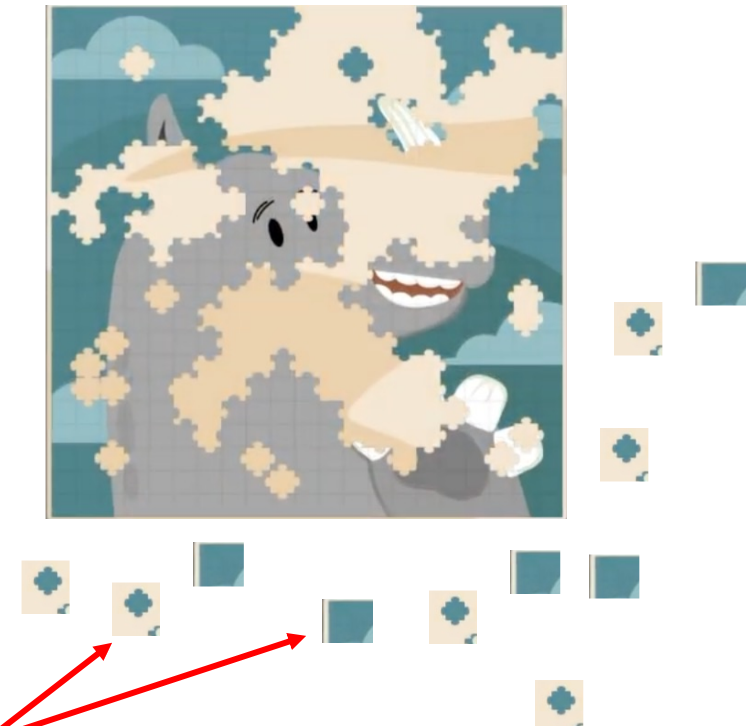
reads still lack information about their order and location

(which part of genome they are originated from)



Solving the Puzzle

Reference genome



Reads

https://www.pacb.com/smrt-science/smrt-sequencing/hifi-reads-for-highly-accurate-long-read-sequencing/

HTS Sequencing Output

Small pieces of a puzzle
short reads (Illumina)

Large pieces of a puzzle
long reads (ONT & PacBio)

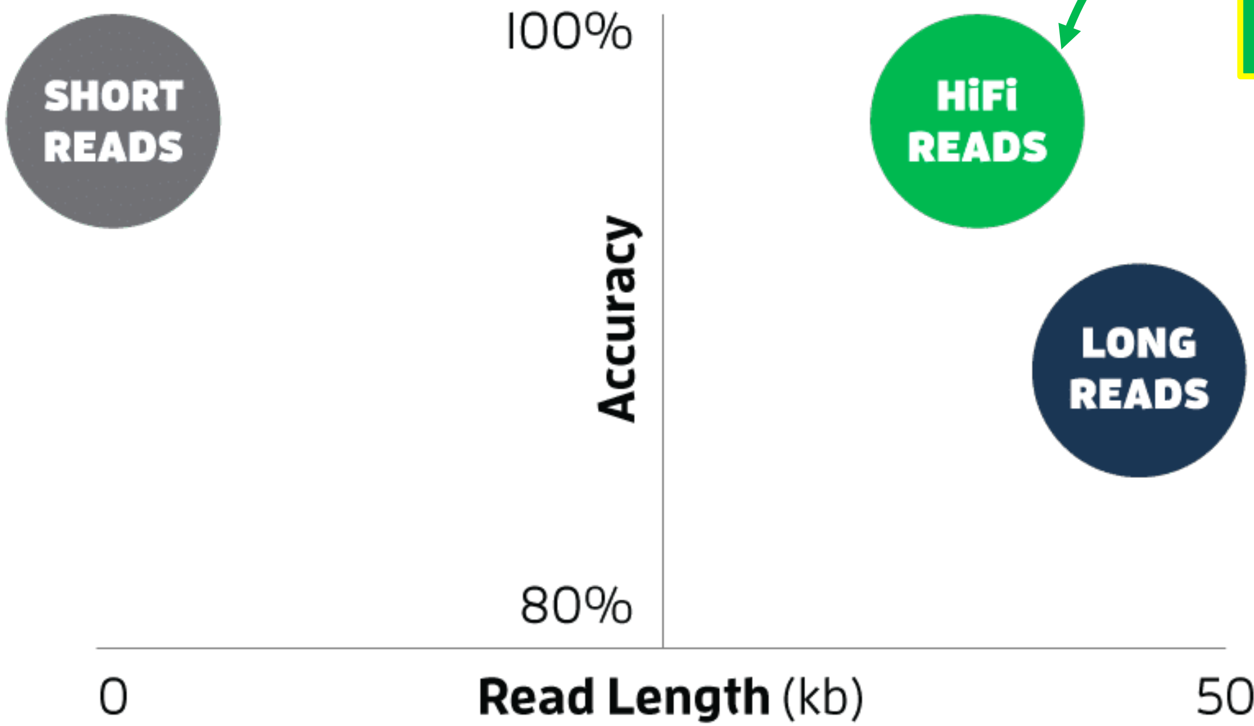
Which sequencing technology is the best?

- 100-300 bp
- low error rate (~0.1%)

- 500-2M bp
- high error rate (~15%)

https://www.pacb.com/smrt-science/smrt-sequencing/hifi-reads-for-highly-accurate-long-read-sequencing/

HiFi Reads (PacBio)



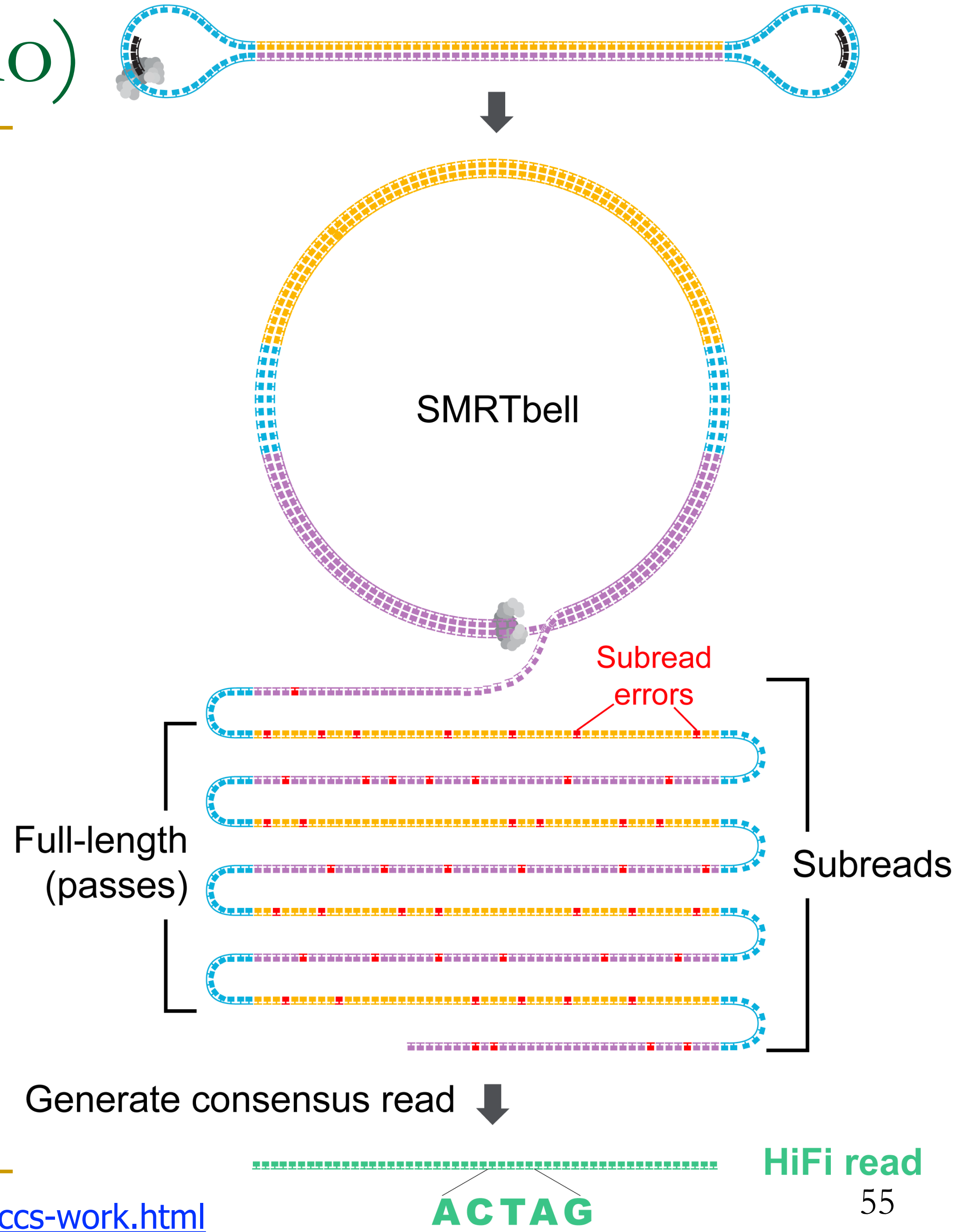
Long: 10-20 kb
Accurate: 99.8%

But still very expensive!

Wenger+, "[Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome]()", *Nature Biotechnology*, 2019

https://labs.wsu.edu/genomicscore/illumina-sequencing/
https://pacbio.gs.washington.edu/

HiFi Reads (PacBio)



Changes in sequencing technologies can render some read mapping algorithms irrelevant

Read Mapping in 111 pages!

In-depth analysis of 107 read mappers (1988-2020)

Mohammed Alser, Jeremy Rotman, Dhrithi Deshpande, Kodi Taraszka, Huwenbo Shi, Pelin Icer Baykal, Harry Taegyun Yang, Victor Xue, Sergey Knyazev, Benjamin D. Singer, Brunilda Balliu, David Koslicki, Pavel Skums, Alex Zelikovsky, Can Alkan, Onur Mutlu, Serghei Mangul
"Technology dictates algorithms: Recent developments in read alignment"
Genome Biology, 2021
[Source code]

Alser *et al. Genome Biology* (2021) 22:249
https://doi.org/10.1186/s13059-021-02443-7

Genome Biology

REVIEW Open Access

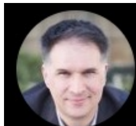
Technology dictates algorithms: recent developments in read alignment

Mohammed Alser[1,2,3†], Jeremy Rotman[4†], Dhrithi Deshpande[5], Kodi Taraszka[4], Huwenbo Shi[6,7], Pelin Icer Baykal[8], Harry Taegyun Yang[4,9], Victor Xue[4], Sergey Knyazev[8], Benjamin D. Singer[10,11,12], Brunilda Balliu[13], David Koslicki[14,15,16], Pavel Skums[8], Alex Zelikovsky[8,17], Can Alkan[2,18], Onur Mutlu[1,2,3†] and Serghei Mangul[5*†]

Feedback From Our Community!

James Ferguson
@Psy_Fer_

This is awesome! I've got my evening reading sorted.



Stéphane Le Crom
@slecrom

Very complete article on the evolution of read alignment algorithms. #NGS #genomics

Svetlana Gorokhova
@SGorokhova

An impressive overview of read alignment methods over the last three decades

BContrerasMoreira @BrunoContrerasM · Sep 10 ···
Replying to @mealser @GenomeBiology and 3 others
Buen hilo de repaso sobre la evolución de los algoritmos de alineamiento de secuencias a medida que ha mejorado la tecnología de secuenciación

https://twitter.com/mealser/status/1435223377644503040

Looking forward,

Will we be able to read

the entire genome sequence?

P&S Mobile Genomics

Lecture 3:
Introduction to Sequencing

Dr. Mohammed Alser

@mealser

ETH Zurich

Fall 2022

31 October 2022

SAFARI

ETH zürich